Genome-Wide Association Studies of malaria susceptibility and resistance: progress, pitfalls and prospects

Delesa Damena¹, Awany Denis, Lemu Golassa², Emile R. Chimusa¹

¹Division of Human Genetics, Department of Pathology, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Private Bag, Rondebosch, 7700, Cape Town, South Africa.

²Addis Ababa University, Aklilu Lema Institute of Pathobiology
PO.box 1176
Addis Ababa, Ethiopia

damenadelesa@gmail.com ; emile.chimusa@uct.ac.za

Abstract

P. falciparum malaria is still among the leading causes of child mortality in sub-Saharan Africa; killing hundreds of thousands of children each year. Malaria has been recognized as one of the prominent evolutionary selective forces of human genome that led to the emergence of multiple host protective polymorphisms associated with minimizing the risk of developing severe malaria in endemic areas. A comprehensive understanding of the genetic bases of malaria resistance can shed light to the molecular mechanisms of host-parasite interactions that can potentially pave ways to the development of new therapeutics and vaccines. Genome-wide association studies (GWASs) have recently been implemented in malaria endemic areas and identified a number of novel association genetic variants. Despite this success, only few variants did replicate across the studies and the underlying biology is yet to be understood for the majority of the novel variants. Besides, there are several open questions around heritability, polygenic effects, epistatic interactions, genetic correlations and associated molecular pathways among others. In this review, we first assess the progress and pitfalls of malaria susceptibility GWASs. We then, provide an overview of the current progress in post-GWAS approaches and discuss how these approaches can potentially be implemented in malaria susceptibility GWASs to extract further functional information. We conclude by highlighting the importance of multi-step and multi-dimensional integrative studies for unravelling the genetic basis of malaria susceptibility and resistance at systems biology level.

Key words: Genome-wide association study, Malaria, *P. falciparum*, susceptibility, heritability,

Introduction

Plasmodium falciparum, the causative agent of the severe malaria, has been infecting humans for at least 5000-10,000 years following the advent and expansions of agriculture¹⁻³. Malaria still poses a huge social, economic and health problems in several low-income countries, particularly in sub Saharan Africa^{4,5}. *P. falciparum* infects millions and kills hundreds of thousands of African children each year. However, this constitutes only small proportion (1%) of the populations in endemic areas in which the infections progress to severe malaria such as profound anemia or cerebral malaria^{6,7}. The genetic basis of such variations in malaria disease severity in endemic areas is mainly unclear as the well-known Mendelian variants only explain 2.5% of heritability of malaria protections⁸.

Comprehensive understanding of the genetic basis of resistance and susceptibility to severe malaria is crucial to understand molecular mechanisms of host-parasite interactions that can inform the development of effective therapeutics, vaccination, diagnostics and risk prediction strategies^{8,9}. In a broader context, the findings from malaria susceptibility studies may also provide new insight to human physiology and genetic susceptibility to other common diseases because ‘malaria is one of the strongest evolutionary forces that shape human genome polymorphisms’⁹.

To better understand the underlying genetics, GWASs have recently been implemented in malaria endemic areas and replicated the well-known variants including *HbS* and *ABO* blood group¹⁰⁻¹³. Despite this success, only few novel variants were identified; suggesting that some of the true association signals might have been attenuated. Which may be resulted from several confounding factors such as: the high genetic diversity of population in malaria-endemic areas, differences in frequencies and effect sizes of protective alleles across populations and the inherent limitations of GWAS approaches among others.

One interesting observation in these studies is the identification of several association signals distributed across the genome that didn’t pass GWAS significance threshold; suggesting the existence of polygenic effects¹⁰⁻¹³. This raises several key questions including 1) What is the genetic architecture of malaria protection? 2) What is the contribution of polygenic effects? 2) What is the proportion of heritability explained by cell-types, chromosomes, functional groups and molecular pathways? 3) What is the extent and pattern of epistasis and pleiotropy at genome wide scale?

Here we review the current status of malaria susceptibility GWASs and provide guidance to future research directions. We begin by assessing the progress, pitfalls and opportunity of malaria susceptibility GWASs. We then provide an overview of the recent progresses in post-GWAS approaches and discuss how these methods can be implemented in Malaria

susceptibility studies to better understand the underlying biology. We conclude by discussing on research areas where further works are needed in light of the global malaria eradication efforts.

Malaria susceptibility GWASs: progress

The GWAS approach is continuing to provide unprecedented successes in genetic studies of wide ranges of complex-traits and common diseases^{14,15,16}. In malaria susceptibility studies, GWAS approach was initially planned to address the acute limitations of the conventional studies, approaches such as candidate gene and linkage analysis that have been implemented in malaria research for more than two decades^{1,17}. The candidate gene-based studies^{18–20} and the family based linkage studies^{21,22} have identified several genomic loci associated with malaria susceptibility/resistance. However, the majority of the findings were discordant and failed to be replicated in different populations; mainly because of their limited power¹. Thus, in view of understanding the genetic basis of malaria susceptibility and resistance at genome wide scale, a global partnership of malaria researchers, named as Malaria Genomic Epidemiology Network (MalariaGEN)¹⁷, has been established and successfully conducted multi-center-scale GWASs^{10–13}.

The GWASs have replicated the well-known variants including *HbS* and *ABO* blood groups which reinforce the importance of erythrocyte variants for protection against severe malaria. Besides, the GWASs have identified novel association variants (Table 1) that may directly or indirectly influence the disease outcome. The epidemiology and biology of the well-known variants were reviewed elsewhere¹. Below we characterize the novel malaria susceptibility genetic variants identified by GWASs. We first discuss the biology of two variants such as *ATP2B4* and cluster of the *glycophorin genes (GYPA/B/E)* that were convincingly associated and independently replicated across malaria endemic populations and extend our discussion to other potential association signals which didn't replicate and/or did not reach the GWAS significance threshold.

Novel association variants

ATP2B4

The association of *ATP2B4* gene with severe malaria was first identified by Timmann *et al.*¹¹ in Ghanaian populations. In this study, several SNPs within the *ATP2B4* gene, located on

chromosome 1q32 showed significant associations. The protective role of the variants of *ATP2B4* gene was subsequently reported in other studies^{12,23}. *ATP2B4* encodes a ubiquitous plasma membrane calcium-transporting *ATPase4* (*PMCA4b*)²⁴. *PMCA4b* is widely expressed in the majority of human tissues and is the main transporter of Ca^{2+} in erythrocyte membrane²⁴. A recent study showed that individuals who have a previously unrecognized haplotype named as ‘haplotype 1’ in their *ATP2B4* gene exhibit a reduced *PMCA4b* expression level²⁵. Several SNPs implicated in malaria protection are localized in this haplotype. Timmann *et al*¹¹, proposed that the disturbance of intracellular Ca^{2+} homeostasis caused by impaired *PMCA4b* regulation might play an important role in interfering with pathological effects of malaria in different tissues. The disturbance of intracellular Ca^{2+} homeostasis in RBCs might impair the parasite development and reproduction in a parasitophorous vacuole membrane (PVM)²⁶. The dis-regulation of intracellular Ca^{2+} in other tissues such as placenta, platelets and brain might modify the adhesion phenotypes and hence minimize adhesion-associated pathologies²⁷.

Protective mechanisms of *ATP2B4* variants might also be associated with impaired regulations of nitric oxide (NO), one of the important molecules in malaria pathogenesis and protections²⁸. Increased level of NO was implicated in protection against cerebral malaria while it increases the risk of uncomplicated malaria²⁹. In neuronal cells, *PMCA4* modulates the activity of neuronal nitric oxide synthesis (nNOS) which in turn regulates NO bioavailability in calcium dependent fashion³⁰. Thus, the reduced production of *PMCA4* might affect the activities of the *nitric oxide synthesis* enzyme in the brain which might affect the concentration of NO and hence protection against severe malaria.

Cluster of the 3 glycophorin genes (GYPA/B/E)

The largest multi-center malaria susceptibility GWAS which included eleven populations was conducted by Band *et al*.¹². In this study, 34 genomic regions containing potential susceptibility loci for severe malaria were identified. Among which, a strong signal was observed at locus between *FREM3* gene and cluster of 3 *glycophorin genes* (*GYPA/B/E*) on Chromosome 4. A haplotype (at SNP *rs184895969*) within this region was reported to reduce the risk of developing severe malaria by about 40% and is common in Kenyan populations with allele frequency of about 10%¹².

In addition to the signals near glycophorin genes, six other putative loci that play key roles in membrane biology were identified in this study. A subsequent study in the same

populations identified a large number of copy number variants which are characterized by deletion, duplications and hybrid structures in *GYPA* and *GYPB* genes³¹. Of which a distinct variant called *DUP4* was reported to reduce the risk of severe malaria by about 40% in eastern African (Kenya) populations. Further characterization showed that, this variant is composed of complex *GYPB-A* hybrid and encode *Dantu* antigen in MNS blood group system³¹. The association of this region with severe malaria was supported by another recent study in Tanzanian populations³². The glycophorin gene cluster, *GYPA* and *GYPB* encode the MNS blood group system and are known to be receptors for *P.falciparum* during RBC invasion³³. *GYPA* and *GYPB* serve as an erythrocyte membrane receptor for *EBA-175* and *EBL-1* proteins of the parasite respectively³⁴. This genomic region is also known to be under an ancient selective pressure resulted from host-pathogen arm races between *P.falciparum* and humans³⁵. Further functional analysis is required to better understand how these variants affect the invasion and/or development of the parasites in erythrocytes and convey protection against severe malaria.

Notable novel association variants

SCO1 and DDC

Notable association signals were identified by the first malaria susceptibility GWAS conducted in Gambian population¹⁰. The first lead SNP (*rs6503319*) is located close to *SCO1* (*synthesis of cytochrome c oxidase*) gene on chromosome 17p13. *SCO1* is a multi-functional signaling protein which plays an essential role in *mitochondrial cytochrome c oxidase (COX)* copper delivery pathways³⁶. *COX* catalyzes electron transfer from reduced cytochrome c to oxygen and is abundantly expressed in muscles, brain and liver³⁶. Deficiency of *COX* caused by mutations in *SCO1* gene can lead to respiratory distress and severe metabolic acidosis³⁷ which are also the major complications during cerebral malaria³⁸. Further studies are needed to understand how the variants in *SCO1* gene are associated with the pathological pathways of cerebral malaria.

The second notable association signal identified in this study was *Dihydroxyphenyl-alanine decarboxylase (DDC)* on Chromosome 7p12.2. A recent study in Tanzanian populations replicated the association of *DDC* variants with cerebral malaria³². *DDC* gene encodes *Aromatic-L-amino-acid decarboxylase* enzyme which is involved in biosynthesis of

neurotransmitters such as dopamine and serotonin³⁹. *DDC* is an essential enzyme for brain and nervous developments and its deficiency is associated with reduced cognitive functions³⁹. *DDC* is involved in cellular immunity and contributes in protection against parasitic disease in invertebrates⁴⁰. Furthermore, mutations in *DDC* gene was reported to be associated with refractoriness of *Anopheles gambiae* mosquito against *P.falciparum* parasites⁴¹

MARVELD3

In addition to the *ATP2B4*, Timmann *et al.*¹¹ identified an association SNP (*rs2334880*) on chromosome 16p 22.2 which is linked to *MARVELD3*. However, this association has not been replicated in other studies. *MARVELD3* is one of the components of tight junction proteins in several epithelial and endothelial tissues and is expressed as two alternative spliced variants⁴². These proteins are involved in assembly, development, maintenances and regulations of tight junction. Tight junctions play a major role in intracellular adhesions and involved in sub-cellular signaling mechanisms⁴².

IL-12 receptors and IL-23 receptors

The most recent malaria susceptibility GWAS was conducted in Tanzanian population¹³. In this study, notable associations signals were identified in immune pathways including in interleukin receptors (*IL-23R* and *IL-12RBR2*), in ketch-like proteins (*KLHL3*) and Human Leucocyte Antigen (HLA) regions. *Interleukin-12* is formed from a heterodimer of *IL12B* (*ILp40 subunit*) and *IL-12A* (*ILp35 subunit*)⁴³. *IL-12* plays a vital role in stimulating cell-mediated immune response against intra-cellular pathogens through binding to high affinity *IL-12RB1* and *IL-12RBR2* receptor complexes. It promotes the development of *T-helper cells* (*Th1*) and enhances the production of *INF-γ*, which are known to mediate the clearance of intracellular pathogens⁴³. In malaria, *IL-12* has been implicated in mediating the protective immunity both in experimental animals and humans⁴⁴. *IL-23* is an important pro-inflammatory cytokine that shares p40 subunits with *IL12*⁴⁵. It induces the differentiation of naive *CD4 T*-cells to *IL-17* which plays key roles in pathogenesis of autoimmune diseases⁴⁶. HLA is encoded by the Major Histocompatibility Complex (MHC), the most polymorphic genes known in human genome. The diversity of MHC is believed to be driven by selection pressure from infectious pathogens and known to be associated with the risk of several infectious diseases⁴⁷. HLA variants such as HLA class I antigen (*HLA-Bw53*) and HLA

class II variant (*DRB1*1302-DQB1*0501*) were reported to confer protections against severe malaria in Gambian populations⁴⁷. HLA class I antigen is expressed by liver cells suggesting that T cells (CTL) responses might efficiently act against the liver stage of malaria parasite in individuals with *HLA-Bw53*⁴⁷. Similarly, individuals with *DRB1*1302-DQB1*0501* variant might possess efficient antigen presentation mechanism that can lead to rapid clearance of blood stage parasite⁴⁷.

Variants in immune pathways are of great interest because of their potential to inform the development of effective malaria vaccines¹. The current study is interesting in that several putative variants in immune pathways were identified. However, the power of this study is limited because of relatively smaller sample size and weak significance threshold used to interpret the findings. Therefore, further studies with higher detective power are needed to consolidate the findings.

Table.1 Summary of genetic variants associated with resistance and susceptibility to severe Malaria identified by GWASs

Genomic regions containing the association variants				Major studies													
				Jallow <i>et.al</i> Pop: Gambian ST: GWAS			Timmann <i>et al.</i> Pop: Ghanaian ST: GWAS		Band et al. Pop: African (11 countries) ST: GWAS		Ravenhall <i>et al.</i> Pop: Tanzanian St =GWAS		Manjurano <i>et al</i> ST: Candidate gene		Rocket <i>et al</i> ST: Candidate gene		
Nearest gene name	Ch	Position	SNP ID Ref.allele/alt.allele	MOI	OR	p-value	OR	P-value	OR	p-value	OR	p-value	OR	P-value	OR	P-value	
<i>ATP2B4</i>	01	203658471	rs4951377 (A/G)	DO	-	-	-	-		3.1x10 ⁻⁹	-	-	-	-	-	-	
		203654024	rs 10900585(T/G)	AD	-	-	0.61	1.9x 10 ⁻¹⁰	-	-	-	-	-	-	0.75	1.7x10 ⁻⁹	
		203660781	rs4951074(G/A)	AD	-	-	0.62	1.3x 10 ⁻⁹	-	-	-	-	-	-	-	-	
<i>ABO</i>	09	13 6132908	Chr9:136132908(T/TC)	RE	-	-	-	-	-	1.7 x 10 ⁻¹³	-	-	-	-	-	-	
		13 613290	rs8176719(delG/G)	DO	-	-	1.48	4.3 x 10 ⁻²¹	-	-	-	-	-	-	0.74	5x10 ⁻²³	
<i>HBB</i>	11	8															
		5367231	rs186699529(C/A)	DO	-	-	-	-	-	2.9x10 ⁻²⁸	-	-	-	-	-	-	-
		5518156	rs11036238	AD	0.61	3.9x 10 ⁻⁷	-	-	-	-	0.13	8.6 x10 ⁻¹³	-	-	-	-	
		5518156	rs372091 (C/T)	AD	-	-	0.46	5.6 x 10 ⁻¹⁴	-	-	-	-	-	-	-	-	
<i>GYP A/B/E and FREM3</i>	04	5,227,002	rs334(A/T)	AD	-	-	-	-	-	-	0.07	8.6 x10 ⁻¹³	0.056	4x x10 ⁻¹⁸	0.14	1.6x 10 ⁻²²⁵	
		143,777,125	rs184895969(A/C)	DO	-	-	-	-	0.67	9.5x10 ⁻¹¹	-	-	0.045	1.9x 10 ⁻³	-	-	
<i>MARVELD3 DDC</i>	16 7	71653637	rs2334880 (T/C)	AD	-	-	1.19	1.9 x 10 ⁻⁶	-	-	-	-	-	-	-	-	
			rs10249420(C/G)	AD	0.69	6.8x10 ⁻⁵	-	-	-	-	-	.605	2.3x10 ⁻³	-	-		
			rs1451375	DO	0.75	6.1x10 ⁻⁶	-	-	-	-	-	-	-	-	-		
<i>SOC1</i>	17		rs65033119	AD	1.21	7.2x10 ⁻⁷											
<i>IL23R, IL12RB</i>	01	67731614	rs6682413	RE	-	-	-	-	-	-	0.48	8x10 ⁻⁷	-	-	-	-	

ST: study type; Pop: population; MOI: mode of inheritance; OR: odd-ratio; AD: additive; DO: dominance; RE: recessive

Challenges of malaria susceptibility GWASs

Despite its promising progress, GWAS approach has several limitations including 1) Weak performances in genetically diverse populations⁸, lack of translation of associated loci into suitable biological hypotheses¹⁶, 3) the well-known problem of missing heritability¹⁵, 4) the lack of understanding of how multiple modestly associated loci within genes interact to influence a phenotype⁴⁸, 5) inefficiency in distinguishing between inflation from bias (cryptic relatedness and population stratification) and true signal from polygenicity⁴⁹, 6) the imperfection of asymptotic distribution of current mixed model association or logistic regression in the specific case of low-frequency variants⁵⁰. The discussion on limitations of GWAS approaches is beyond the scope of this review. Here we focus on the major challenges of malaria susceptibility GWASs and highlight recent positive progresses.

Dealing with genetic diversity of African populations

There is high level of genetic diversity and weak linkage disequilibrium (LD) in Africans compared to non-African populations^{51,52} which is believed to be due to adaptation to diverse climate, diet and response to infectious diseases such as malaria and tuberculosis^{51,53}. These distinct genetic characteristics created major setbacks to GWASs in African populations⁸. The fundamental rationale of the GWAS is a ‘common disease common variant’ hypothesis which predicts that common diseases are influenced by common genetic variants that are prevalent in higher frequencies in the populations¹⁶. To capture optimum amount of the common variations in human genome, commercial arrays that constitute representative SNPs (tag SNPs) are utilized. The selection of tag SNPs (subset of common SNPs) for genotyping chips is based on LD information derived from reference panel in which African populations are not fairly represented⁵⁴.

Consequently, the African GWASs might have missed several true association signals. For instance, in the first malaria GWAS¹⁰, *HbS* locus, a well-known variant conferring resistance to severe malaria demonstrated a weak signal (p-values $\sim 10^{-7}$) because of the weak LD between causal variants and the SNPs that were genotyped. After which authors performed targeted sequencing of the locus, undertook multipoint imputation, used proper reference panel and dramatically improved the signal to p-value $\sim 10^{-14}$. It was estimated that a GWAS of 0.6 million SNPs based on HapMap phase 1 dataset in European population

has an equivalent power to the chips with 1.5 million SNPs in African populations⁵⁴. However, the coverage of commercial genotyping arrays has been enormously improved to be able to capture the genetic diversities among global populations following the recent technological advances and availability of diverse reference data sets. For instance, Omni microarrays based GWASs were shown to be powerful in studying African populations⁵². Considerable progresses have been made in bringing African population genomes to the public domain. For instance, different ethnic groups from countries such as: Nigeria, Kenya, Gambia, Ghana and Malawi were included in the International Hapmap 3 Consortium⁵⁵ and 1000 Genomics project consortium that provided a sufficiently deeper sequence for all the major human ancestries⁵⁶. A more diverse set of African population that includes genomic data of 18 ethno-linguistic groups was recently published by The African Genome Variation Project (AGVP)⁵². The Consortium on Asthma among African ancestry Population in America (CAAPA)⁵⁷ is also an additional resource that can facilitate genomic analysis in populations of African ancestry. The improvement in reference panels have facilitated imputation-based studies in African populations. For instance, Band *et al.*⁹ showed the feasibility of multi-point imputation based meta-analysis in for malaria GWASs using HapMap3 haplotype panel. However, imputation accuracy was relatively lower compared to that of European populations. Another study showed a substantial improvement of imputation accuracy by using the more diversified AGVP WGS reference panel⁵². The reference dataset is expected to grow further and accelerate genomic research by including wide range of haplotype diversity in African populations.

Sample size

In GWAS, a stringent p-value (1×10^{-8}) is usually needed to declare evidences of genuine associations to minimize false discovery rate that can arise from multiple testing.⁵⁸ Thus, very large sample size is required to achieve genome-wide significance threshold particularly for loci with modest effect sizes. The required sample size is even much higher for studies in population of African ancestry because of the higher genetic diversity. However, the sample sizes of GWASs in African populations are generally small compared to non-Africans⁵⁹.

In malaria susceptibility GWASs, the success rate of identifying significant associations seems to be a function of sample sizes. in the first Malaria GWAS¹⁰ that involved 2500 children, none of the association signals passed the GWAS significance threshold; whereas in the subsequent GWASs that involved relatively larger sample sizes^{11,12}, novel

association variants were convincingly identified. To this effect, MalariaGen consortium has increased the sample size and included more study populations in its recent GWASs¹². However, one of the challenges of the MalariaGEN dataset is that the sample sizes obtained from each study population is limited; making it difficult to undertake separate GWASs for specific geographic areas. Allelic heterogeneity of the malaria protective variants is very common in endemic areas. For instance, *HbC*, one of the malaria protective alleles, is common in some parts of west Africa such as Burkina Faso, Ghana, Togo and Benin while absent in other west African countries such as Cameroon and Chad⁶⁰. The sickle cell allele, *HbS*, is known to have different haplotype structure and effect sizes in different regions across sub-Saharan Africa⁵². A recently identified glycophorin variants are common in east Africa and absent in some populations in west Africa¹². This may suggest that the majority of variants driving the malaria clinical severity are heterogeneous; pressing the need of increasing sample sizes for each study populations.

Genetic architecture of malaria susceptibility/resistance

The performance of GWASs is dependent on the genetic architecture of the disease/trait under investigation. For the majority of complex diseases/traits, the GWAS variants identified thus far, only explain a very small proportion of heritability; a phenomenon commonly termed as ‘missing’ heritability¹⁵. There have been different explanations for the ‘missing’ heritability including common disease rare variant hypothesis⁶¹, none-additive components, primary epistasis^{62,63} and polygenic genetic architecture¹⁶.

One of the challenges of malaria GWASs is that we don’t know much about the genetic architecture of malaria protection trait. First, as one of the prominent evolutionary selective forces, the majority of malaria protective alleles might have evolved under positive selection and might potentially be balanced by other forces⁶⁴. In this case, the protective variants are expected to have large effect sizes and higher allele frequencies which can be detected by the conventional GWAS approaches; provided that proper reference panel and genotyping array are used⁶⁴. Second, similar to the genetic architecture of other infectious diseases⁶⁵, malaria protection trait might largely be attributed to few rare variants of large effect sizes. In this, the GWASs are underpowered as rare variants might not be in LD with common variants and can’t not be ‘tagged’ by SNPs chip. Third, malaria protection trait might be mainly under pyogenic and epistatic control as suggested by several authors^{1,7,66} which the conventional GWAS approach can’t capture. Here, we focus on polygenic genetic architecture and epistasis which have been implicated for the ‘missing’ heritability of malaria susceptibility GWASs.

Polygenic architecture and epistasis: Presenting the absent in the current malaria GWASs

Polygenic genetic architecture

Polygenic view of genetic architecture is gaining ground in genetic epidemiological studies and widely implicated for the ‘missing’ of heritability in GWAS analysis⁶⁷. The rationale behind polygenic inheritance is that complex-traits/diseases are influenced by multiple variants with modest effects that are too small to pass the stringent genome wide significance threshold⁶⁸. In standard GWAS analysis, ‘Genomic control’ (GC) method is applied as a quality-control measure to minimize spurious associations that can be caused by population structures such as population stratification and cryptic relatedness.

However, a slight inflation of the test statistics which cannot be corrected by GC was initially observed across the genome in a Schizophrenia GWAS⁶⁹. Subsequently, this observation has been supported by other studies⁷⁰ and led to the development of a number of statistical tools aiming to capture polygenic signals at genome-wide scale including 1) polygenic scoring method implemented in PLINK software⁷¹; 2) Mixed Linear Models (MLM) such as: GCTA⁷⁰, BOLT-LMM⁷², Bayes R⁷³ and LDAK^{74,75}, PCGC⁷⁶ 3) linkage-disequilibrium (LD) score regression method⁷⁷ among others.

In polygenic scoring method, two GWAS sample sets are used. First, a ‘training sample’ data set is used to estimate the effect size of each SNPs. Then, subset of SNPs with top ranking p-values are selected to calculate polygenic score in an independent dataset called ‘target sample’ for each individual based on the weighted sum of risk alleles at selected SNPs. The detail of this approach is explained in Purcell *et al.*⁶⁹. The MLM approaches, that jointly accounts for fixed and random effects, is becoming a standard method for estimating heritability of complex diseases/traits. Yang *et al.*⁷⁰ introduced an elegant method that builds genetic relationships matrix (GRM) of unrelated individuals by simultaneously fitting all SNPs and estimates the proportion of phenotypic variations explained by genotypic variation using restricted maximum likelihood (REML) method. Variations of MLM approaches with different flavors have subsequently been published and implemented for estimating and stratifying heritability of complex traits/diseases in to chromosomes, functional groups and pathways^{74–76}. The use of unrelated individuals has an advantage of minimizing the bias from shared environments in pedigree studies.

Another approach is LD score regression method⁷⁷. The rationale of this approach is that there are three sources of variations in GWAS estimates such as true signals, population structure and estimation errors. The LD score regression effectively decomposes these components by exploiting the LD strength of a given SNP with other SNPs. The marginal association is higher for SNPs that are in LD with several variants. The detailed description of the polygenic analysis methods is reviewed in⁶⁷.

Polygenic contributions in malaria susceptibility/resistance

Co-evolution of host-pathogen model predicts that multiple host loci are involved in resistance/susceptibility to infectious diseases due to the complex interactions between the multi-locus parasite genotype and the corresponding defense from the host-genome^{63,78}. Indeed, malaria might have left multiple genetic variants; the majority of which have effects too small to be detected by the standard GWASs. The existence of polygenic inheritance in malaria protection was predicted by several authors^{1, 7} and supported by GWASs. For instance, the largest malaria susceptibility GWAS so far, identified 34 regions of the genome containing variants with evidence of associations¹². Earlier GWAS in Ghanaian population identified 40 genomic regions containing 102 SNPs with evidences of association in the discovery phase of the study¹¹. The recent GWAS in Tanzanian populations¹³ identified 2322 SNPs at several genomic regions by relaxing the significance threshold to 1×10^{-5} . Thus, implementation of polygenic analytic methods in malaria studies may potentially shed more light to the underlying biology. For instance, heritability can be estimated and partitioned in to different cell-types and functional groups and molecular pathways which enable to localize causal variants. Furthermore, these approaches can be extended to explore genetic correlation between susceptibility to malaria and to other infectious diseases. The existence of shared genetic basis between infectious diseases susceptibility/protection is well-documented⁷⁹. However, the extent and pattern of genetic correlation has not been systematically investigated at genome-wide scale; partly because of inadequate GWAS data for infectious diseases. Such studies can potentially provide clues to common molecular processes between resistance/susceptibility to infectious diseases that will have practical importance such as: designing multi-purpose vaccine and genetic risk prediction strategies.

Heritability of severe malaria in Gambian population

To figure out how polygenic analysis can be implemented in malaria susceptibility, we accessed the Gambian malaria susceptibility GWAS dataset from European Phenome Genome Archive (EGA) through data application procedure and estimated heritability of malaria susceptibility using MLM approaches. The Gambian GWAS data is the largest MalariaGen dataset obtained from a single country comprised of 4920 samples (2429 cases and 2491 controls) and 1.6 million SNPs that passed GWAS quality control (QC). We first excluded the known malaria susceptibility associated loci and performed stepwise extra QC filtering. Specifically, we focused on sample relatedness, SNP missingness proportion and SNP differential missing proportion which are well-known to affect the accuracy of heritability estimation⁷⁵. We then estimated the heritability using GCTA model for different QC thresholds by including 10 principal components (PCs) as fixed effects to account for population structure. As expected, the estimation is highly unstable when less stringent QC thresholds is applied (varying from 37.8% to 20.1%) as shown in Table 2. However, when more stringent QC (Relatedness threshold (5%), SNP differential missingness proportion ($p \leq 1 \times 10^{-3}$) and SNPs missing proportion of ($p > 0.02$)) is applied, the estimation became stable (~20.1%, SE = .07). Neither the inclusion of more PCs (15, 20) as fixed effects nor SNP phasing further brought down the estimate. Using the same stringent QC threshold, the estimation was approximately the same for Mandinka ethnic group (~24.3%, SE=0.2). We couldn't estimate for other ethnic groups because of smaller sample sizes. Furthermore, the use of PCGC model which is designed for case/control approximately showed the same estimate (19.8%, SE=.07). However, the LDAK model completely showed different estimate (41.0%).

Encouraged by the fairly stable estimate, we partitioned the heritability of severe Malaria heritability in to intergenic and genic regions using GCTA model as shown in table (3). The proportion of heritability explained by a SNP in gene within 10kb boundary is on average twice as the heritability explained by a SNP in intergenic regions based on 10kb boundaries. Larger sample size is needed to further partition the heritability explained by cell types and functional groups.

Although our heritability estimation is fairly stable when stringent QC is implemented, care should be taken in interpreting these results: First, all polygenic methods perform better in less structured data obtained from homogenous populations than the MalariaGen dataset which is comprised of diverse populations spanning most of the Malaria endemic belt in Africa. Second, the methods are designed and perform well in highly polygenic

traits/diseases in which effects of each variant is mostly modest. However, a significant proportion of malaria protection trait might be attributed to rare variants of large effect sizes that might not be in LD with common variants and can't not be 'tagged' by SNPs chip. The contribution of such variants will not be accounted for in the estimation. Third, subtle population structure that cannot be corrected by conventional methods such as unmatched case/control can potentially create systematic biases to the estimate. Fifth, the controversy between the performances of different methods is still continuing and there is no gold standard method against which the accuracy of these models can be evaluate

Table 2. SNP-heritability of severe malaria susceptibility/resistance of Gambian population for different basic quality threshold using MLM

Population	Sample relatedness-threshold	SNP missingness-proportion	SNP differential-missingness Proportion	Prevalence	Covariate principal-components	No. Samples	No. Snps	GCTA h2(% SE)	PCGC h2(SE)	LDAC
Gambia	-	5%	-	1%	10	4920	1627656	37.8(.06).		
	5%	5%	1x 10 ⁻¹⁰	1%	10	4128	1627656	30.5(.07)		
	5%	5%	1x10 ⁻⁵	1%	10	4128	1607610	28.7(.07)		
	5%	5%	1x 10 ⁻³	1%	10	4128	1570344	25.1(.07)		
	5%	2%	1x 10 ⁻³	1%	10	4128	1486554	20.1(.07)	19.8(.07)	41.0(.08)
	5%	2%	1x 10 ⁻³	1%	15	4128	1486554	22.5(.07)		
	5%	2%	1x10 ⁻³	1%	20	4128	1486554	19.5(.07)		
	Phased	-	-	1%	10	4128	1627656	20.4(.08)		
Mandinka	5%	2%	1x10 ⁻³	1%	10	1281	1486554	24.2(0.2)		

GCTA: Genome Complex Trait Analysis; PCGC: Phenotype Correlation Genotype Correlation regression

Table 3. Partitioning of heritability in to genic and intergenic genomic regions based on 10 kb boundaries using UCSC annotation data bases

	10 Kb boundaries of genes	
	Genic	Inter-genic
No. of SNPs	827928	685896
h ² (%SE)	16.3 (.06)	6.4 (.05)
h ² per SNP	1.9x10 ⁻⁵	9x10 ⁻⁶

Epistasis

Epistasis is becoming one of the hot research topics in genetic epidemiological studies in the last few years because of the fact that none additive genetic variations are shown to have significant influence on the phenotype of complex traits/diseases than previously expected⁸⁰. The available statistical approaches and software packages for detection of epistasis at genome-wide scale are reviewed elsewhere^{80,81}. These approaches have been applied in genetic studies of complex diseases such as lupus erythematosus⁸², ankylosing spondylitis⁸³, psoriasis⁸⁴ and unraveled previously unknown epistasis interactions between risk loci which explained a significant proportion of ‘missing’ heritability of the respective diseases.

Epistasis between malaria risk loci have been well documented and implicated as one of the possible reasons for lack of replication of susceptibility variants in different populations and the ‘missing’ heritability. For instance, sickle-cell trait (*HbS*) and α *thalassaemia* were shown to demonstrate negative epistatic interactions such that the protection against severe Malaria offered by *HbS* is reduced when co-inherited with α *thalassaemia*⁸⁵.

A case-control study in Kenyan population also reported that α *thalassaemia* modulates the effects of Haptoglobin (Hp) variants in predicting the risk of severe malaria⁸⁶. In this study, it was shown that the combination of α *thalassaemia* and Hp2-1 variant synergistically increase the protection against severe malaria by about 37%. However, the protective effect is decreased to 13% when α *thalassaemia* is inherited with Hp1-1 and further diminished to neutral (zero) when inherited with Hp2-2. Similarly, in a multi-center case control study, the existence of negative epistasis interaction between *HbC* and *ATP2B* alleles was reported¹⁸.

Another recent case control study of severe malaria in Kenya reported the existence of negative epistasis between a complement receptor called *SL2* and α *thalassaemia* in which the protective effect of *SL2* higher in children with normal α -globin⁸⁷. The extent and pattern of epistatic interaction at genome-wide scale is yet to be explored. In malaria susceptibility GWAS, the priority has been given to a single locus analysis to identify novel risk loci. We expect that, the next step of Malaria susceptibility GWAS will be the considerations of epistasis.

From GWAS to biology: multi-step and multi-dimensional analysis

Fine-mapping and pathway analysis

The ultimate goal of genetic susceptibility studies is to identify the causative variants and understand the underlying biological pathways which can lead to translated medicine such as effective vaccines and therapeutics. However, translating GWAS signals into biological

themes remains an open problem because of the confounding effects from LD between association SNPs, limited knowledge of gene functions and localization of the majority of GWAS hits in non protein coding regulatory regions (regulatory SNPs)^{88–90}. In attempt to address this challenge, several fine mapping strategies have been developed and implemented⁹¹. One such strategy is trans-ethnic fine mapping in which the natural variability of haplotype structure across ethnically diverse populations is used to narrow down the candidate causative variants⁹². The smaller LD and diversity of haplotype structure in African population makes it relatively easier to identify the causal SNPs and target gene/genes through fine mapping approaches⁸. However, the fact that malaria protective alleles are heterogeneous across populations might challenge the application of trans-ethnic fine-mapping approaches in malaria susceptibility studies.

Alternatively, several fine mapping statistical tools have recently been developed following the advances in annotation data bases and improved reference panels. These include Bayesian approach, heuristic approach and penalized regression methods⁹³. The principles, applications, strength and weakness of these methods is reviewed elsewhere^{91,93}. These methods are increasingly playing crucial role in the efforts being made to pinpoint causal variants of complex diseases/traits. For instance, Galarneau *et al.*⁹⁴ identified novel independent association signals by fine-mapping three loci that are known to influence fetal hemoglobin (HbF) levels. The authors sequenced the three loci (*BCL11A*, *β-globin* and *HBS1L-MYB*), undertook dense genotyping, performed step-wise conditional analysis and revealed previously un-recognized SNPs that explain additional genetic variation. Similarly, a recent fine mapping study of HLA region identified several susceptibility loci for multiple infectious diseases⁷⁹. More sophisticated studies that combine statistical and functional fine-mapping strategies have recently been implemented and provided mechanistic insights to the genetic basis of complex diseases⁹⁵.

In addition to fine mapping approaches, pathway and interaction analysis can be another avenue for exploring molecular basis of Malaria susceptibility/resistance. Instead of emphasizing on single-variant analysis, these approaches test the coordinated effects of several variants at systems level using biological information from annotation data basis⁹⁶. Pathway analysis improve study power by integrating cumulative effects of weak association signals and provide functional information by identifying associated sets of genes/proteins⁹⁷. By implementing the pathway analysis approaches, several studies have gained new insights in understanding the genetic basis of complex diseases^{98–100}. The available statistical tools for pathway and interaction analysis is reviewed in¹⁰¹. We

therefore, advocate for the implementation of fine mapping and pathway analysis in malaria susceptibility studies to shed more light to the underlying biology.

Multi-omics approaches

Today, there are significant advances in high throughput technologies that can generate big ‘-omic’ data from all spectrum of molecular biology⁸⁸. The ‘-omics’ studies (Genomics, Epigenomics, Transcriptomics, Proteomics, Metabolomics) are crucial to understand the underpinning biology of complex diseases.

In severe malaria, ‘-omics’ studies have provided important clues about the molecular events that lead to either complications or recovery from diseases. For instance, following the discovery of glycophorin regions by the GWAS, a targeted sequencing-based study³¹ was conducted to characterize variants in this region and identified a novel distinct copy number variant called *DUP4* which reduces the risk of severe Malaria by about 33% . In addition to this, a genome-wide gene expression study was conducted in Kenyan children in which increased expression of genes related to neutrophil activation during malaria infections was shown¹⁰². The authors also observed differential expression of heme- and erythrocytes-related genes in acute malaria patients which reaffirms the importance of erythrocyte-related genes in malaria susceptibility/resistance.

In another host-parasite interaction study, the importance of miRNA in inhibiting parasite growth in erythrocyte was reported¹⁰³. The authors observed translocation of several host RBC miRNAs in to *P. falciparum* parasites, as well as fusion of these human miRNAs with parasite mRNA transcripts to inhibit the translation of enzymes that are vital for the parasite development. Specifically, two micro-RNAs, miRNA-451 and let-7i, were highly enriched in *HbAS* and *HbSS* erythrocytes and these miRNAs along with miR-223 were shown to attenuate the growth of parasite¹⁰³.

However, ‘-omics’ studies are limited to single data-type analysis and lack adequate power to explain the complexity of molecular processes and usually lead to identification of correlations than causations¹⁰⁴. Thus, integrating and analysing multiple ‘-omics’ data enables better understanding of the molecular processes and interactions that give rise to complex diseases/traits. For example, leveraging host microbiome relative abundance data as a second (quantitative) trait, and performing a joint analysis of bivariate phenotypes can increase statistical power by maximizing phenotypic information and inform how the interaction between host genotype with microbiome impacts the phenotype.

Multi-omics approaches aim to integrate big ‘-omics’ data, undertake ‘multi-step’ and ‘multidimensional’ analysis for elucidating complex biological problems¹⁰⁴. Driven by the massive abundance of ‘-omics’ data from wide ranges of biological molecules, multi-omics strategy have recently provided unprecedented successes in complex diseases/trait studies. The current state of art of multi-omics approach and available statistical methods is recently reviewed in Hasin *et al.*¹⁰⁴

Malaria susceptibility/resistance is influenced by several host, parasite and environmental factors as depicted in Figure 1. The protective alleles have independently evolved in different populations being shaped by the co-evolution and interaction between the human genome, the parasite and environment¹. So far, single ‘-omics’ data analysis enabled us to understand some of the factors that are associated with the malaria protective traits. To progress beyond associations and pinpoint the causal pathways, it may require to implement carefully designed, coordinated multi-omics studies that involve human host, the parasite, the environment and possibly mosquito. The current advents of high-throughput technologies in generating massive ‘-omics’ data and their continuously decreasing cost complemented with the availability of statistical tools which able to simultaneously capture millions of data points will lead to the implementations of multi-omics approaches in malaria susceptibility studies.

Figure 1: Schematic representation of the integrative analysis. Systems biology approach which incorporate multiple layers of information from host (multi-omics), environment and parasite genetic factors can potentially lead to the discovery of Malaria protective pathways

Conclusions and perspectives

The ultimate goal of Malaria susceptibility study is to discover a novel causal biological pathway that provide protections against severe Malaria; a fundamental step towards translational medicine such as development of vaccine and new therapeutics that can facilitate the global malaria eradication efforts. To achieve this goal, several study approaches have been implemented at least for the last three decades and successfully identified association variants.

Recently, a number of GWASs have successfully been implemented in malaria endemic areas to better understand the underlying biology. While the well-known variants were replicated, only few novel variants were convincingly identified and their biological functions remains to be understood. Several limiting factors including genetic diversity of population in malaria endemic areas, allelic heterogeneity of protective variants, small sample sizes, lack of proper reference panel and proper genotyping chips might have negatively impacted the malaria GWASs.

Another challenge is that we don't know much about the genetic architecture of malaria protective trait. There are two scenarios in which GWAS approach might fail. First, malaria protective trait might largely be attributed to rare variants of large effect sizes that might not be in LD with common variants and can't not be 'tagged' by SNPs chip. Second, malaria might have left multiple genetic variants distributed across human genome; the majority of which have effects too small to be detected by the standard GWASs^{1,7}. Theoretically, the large sample sizes, dense genotyping chips, use of appropriate reference panels and effective genotype imputation can address the majority of challenges. However, given the resource constraints; especially, in Africa where malaria problem is the greatest, this will likely take several years to achieve.

On the other hand, the recent advances in statistical techniques is enabling to extract useful information from the present-day GWAS sample sizes. For example, a number of statistical approaches have been developed to capture polygenicity in complex diseases. We showed how these methods can potentially be implemented in malaria susceptibility studies and provide useful insights. We believe that further studies with larger sample sizes can elucidate the polygenic effects in malaria protective trait by extending the analysis to genome partitioning, risk prediction and genetic correlations.

Beyond single locus analysis, multi-step and multi-locus analysis including pathways analysis, fine mapping and interaction analysis can potentially be implemented in malaria susceptibility GWASs to gain new insights to the underpinning biology. For instance, pathway analysis can provide important information by analyzing the coordinated effects of several variants at systems level using biological information from annotation data basis. Fine mapping strategies that combine statistical and functional fine-mapping strategies can potentially be implemented to pinpoint the causal variants from the GWAS association signals.

Most importantly, the future direction of malaria susceptibility requires a paradigm shift from single ‘-omics’ to ‘multi-stage’ and ‘multi-dimensional’ integrative functional studies that combines multiple data types from the human host, the parasite, the mosquitoes and the environment. The current biotechnological advances, an ever-increasing annotation data bases and availability of advanced analytical techniques, will eventually lead to feasibility of systems biology studies and revolutionize malaria research.

Acknowledgements

We thank Kwiatkowski’s group from University of Oxford for their constructive comments and assistance. We thank Kirk Rokett and Ellen Leffler for their useful comments. We are very grateful to Gavin Band for his supervision and guidance in heritability analysis and his critical comments. I thank Newton’s fund student transfer scheme for funding me during my stay at University of Oxford. We also thank Abdoulaye Djimde for his follow up.

Funding

DD is a PhD student funded by DELGEME (<http://delgeme.org/>) program. EC is funded by NIH projects

Declaration of Interest

The authors declare that they have no competing interests.

Authors contributions

DD designed and drafted the manuscript; LG and AD revised the manuscript; EC revised the manuscript and supervised the work

References

1. Kwiatkowski, D.P. (2005). How Malaria Has Affected the Human Genome and What Human Genetics Can Teach Us about Malaria. *Am. J. Hum. Genet.* 77, 171–192.
2. Parikh, S., and Rosenthal, P.J. (2008). Human Genetics and Malaria : Relevance for the Design of Clinical Trials. *198*, 1255–1257.
3. Sabeti, P., Usen, S., Farhadian, S., Jallow, M., Doherty, T., Newport, M., Pinder, M., Ward, R., and Kwiatkowski, D. (2002). CD40L association with protection from severe malaria. *Genes Immun.* 3, 286–291.
4. Neghina, R., Iacobiciu, I., Neghina, A.M., and Marincu, I. (2010). Malaria, a Journey in Time: In Search of the Lost Myths and Forgotten Stories. *Am. J. Med. Sci.* 340, 492–498.
5. World Health Organization (2017). World Malaria Day report. WHO, Geneva, Switzerland
6. Marsh, K., Forster, D., Waruiru, C., Mwangi, I., Winstanley, M., Marsh, V., Newton, C., Winstanley, P., Warn, P., Peshu, N., et al. (1995). Indicators of Life-Threatening Malaria in African Children. *N. Engl. J. Med.* 332, 1399–1404.
7. Mackinnon, M.J., Mwangi, T.W., Snow, R.W., Marsh, K., and Williams, T.N. (2005). Heritability of malaria in Africa. *PLoS Med.* 2, 253–1259.
8. Teo, Y.-Y., Small, K.S., and Kwiatkowski, D.P. Methodological challenges of genome-wide association analysis in Africa. *Nat Rev Genet.* 11: 149–160
9. Band, G., Le, Q.S., Jostins, L., Pirinen, M., Kivinen, K., Jallow, M., Sisay-Joof, F., Bojang, K., Pinder, M., Sirugo, G., et al. (2013). Imputation-Based Meta-Analysis of Severe Malaria in Three African Populations. *PLoS Genet* 9: e1003509.
10. Jallow, M., Teo, Y.Y., Small, K.S., Rockett, K.A., Clark, T.G., Kivinen, K., Bojang, K.A., Conway, D.J., Sirugo, G., Sisay-joof, F., et al. (2010). Genomewide and fine-resolution association analysis of malaria in West Africa. *Nat. Genet.* 41, 657–665.
11. Timmann, C., Thye, T., Vens, M., Evans, J., May, J., Ehmen, C., Sievertsen, J., Muntau, B., Ruge, G., Loag, W., et al. (2012). Genome-wide association study indicates two novel resistance loci for severe malaria. *Nature* 489, 443–446
12. Band, G., Rockett, K.A., Spencer, C.C.A., Kwiatkowski, D.P., Band, G., Si Le, Q., Clarke, G.M., Kivinen, K., Leffler, E.M., Rockett, K.A., et al. (2015). A novel locus of resistance to severe malaria in a region of ancient balancing selection. *Nature* 526, 253–257.
13. Ravenhall, M., Campino, S., Sepu, N., Nadjm, B., Mtove, G., Wangai, H., Maxwell, C., Olomi, R., Reyburn, H., Drakeley, C.J., et al. (2018). Novel genetic polymorphisms associated with severe malaria and under selective pressure in North-eastern Tanzania. *PLoS Genet* 14(1): e1007172
14. Visscher, P.M., Brown, M.A., McCarthy, M.I., and Yang, J. (2012). Five Years of GWAS Discovery. *Am. J. Hum. Genet.* 90, 7–24.
15. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 Years of GWAS Discovery : Biology , Function , and Translation. *Am. J. Hum. Genet.* 101, 5–22.
16. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. *Nature* 461, 747–753.
17. Achidi, E.A., Agbenyega, T., Allen, S., Amodu, O., Bojang, K., Conway, D., Corran, P., Deloukas, P., Djimde, A., Dolo, A., et al. (2008). A global network for investigating the genomic epidemiology of malaria. *Nature* 456, 732–737.
18. Malaria Genomic Epidemiology Network, (2014). Reappraisal of known malaria resistance loci in a large multicenter study. *Nat Genet.* 46, 1197–1204
19. Verra, F., Mangano, V.D., and Modiano, D. (2009). Genetics of susceptibility to *Plasmodium falciparum*: From classical malaria resistance genes towards genome-wide association studies. *Parasite Immunol.* 31, 234–253.
20. Hedrick, P. (2011). Population genetics of malaria resistance in humans. *Heredity.* 107/16, 283–304.
21. Flori, L., Sawadogo, S., Esnault, C., Fre, N., Fumoux, F., and Rihet, P. (2003). Linkage of

- mild malaria to the major histocompatibility complex in families living in Burkina Faso. *Human Molecular Genet.* *12*, 375–378.
22. Rihet, P., Traore, Y., Abel, L., Aucan, C., and Fumoux, F. (1998). Malaria in Humans : Plasmodium falciparum Blood Infection Levels Are Linked to Chromosome 5q31-q33. *Am J Hum Genet.* *63*, 498–505.
 23. Bedu-Addo, G., Meese, S., and Mockenhaupt, F.P. (2013). An ATP2B4 polymorphism protects against malaria in pregnancy. *J. Infect. Dis.* *207*, 1600–1603.
 24. Guerini, D., Pan, B., and Carafoli, E. (2003). Expression, purification, and characterization of isoform 1 of the plasma membrane Ca²⁺ pump. Focus on calpain sensitivity. *J. Biol. Chem.* *278*, 38141–38148.
 25. Zámbo, B., Várady, G., Padányi, R., Szabó, E., Németh, A., Langó, T., Enyedi, Á., and Sarkadi, B. (2017). Decreased calcium pump expression in human erythrocytes is connected to a minor haplotype in the ATP2B4 gene. *Cell Calcium* *65*, 73–79.
 26. Gazarini, M.L., Thomas, A.P., Pozzan, T., and Garcia, C.R.S. (2003). Calcium signaling in a low calcium environment: How the intracellular malaria parasite solves the problem. *J. Cell Biol.* *161*, 103–110.
 27. Adams, Y., Kuhnrae, P., Higgins, M.K., Ghumra, A., and Rowe, J.A. (2014). Rosetting Plasmodium falciparum-infected erythrocytes bind to human brain microvascular endothelial cells in vitro, demonstrating a dual adhesion phenotype mediated by distinct P. falciparum erythrocyte membrane protein 1 domains. *Infect. Immun.* *82*, 949–959.
 28. Sobolewski, P., Gramaglia, I., Frangos, J., Intaglietta, M., and Van Der Heyde, H.C. (2005). Nitric oxide bioavailability in malaria. *Trends Parasitol.* *21*, 415–422.
 29. Turner, G.D., Morrison, H., Jones, M., Davis, T.M., Looareesuwan, S., Buley, I.D., Gatter, K.C., Newbold, C.I., Pukritayakamee, S., and Nagachinta, B. (1994). An immunohistochemical study of the pathology of fatal malaria. Evidence for widespread endothelial activation and a potential role for intercellular adhesion molecule-1 in cerebral sequestration. *Am. J. Pathol.* *145*, 1057–1069.
 30. de Jesus Trovoada, M., Martins, M., Mansour, R. Ben, Do Rosário Sambo, M., Fernandes, A.B., Gonçalves, L.A., Borja, A., Moya, R., Almeida, P., Costa, J., et al. (2014). NOS2 variants reveal a dual genetic control of nitric oxide levels, susceptibility to Plasmodium infection, and cerebral malaria. *Infect. Immun.* *82*, 1287–1295.
 31. Leffler, E.M., Band, G., Busby, G.B.J., Kivinen, K., Le, Q.S., Clarke, G.M., Bojang, K.A., Conway, D.J., Jallow, M., Sisay-Joof, F., et al. (2017). Resistance to malaria through structural variation of red blood cell invasion receptors. *Sci.* *356*, 1140–1152.
 32. Manjurano, A., Sepulveda, N., Nadjm, B., Mtove, G., Wangai, H., Maxwell, C., Olomi, R., Reyburn, H., Drakeley, C.J., Riley, E.M., et al. (2015). USP38, FREM3, SDC1, DDC, and LOC727982 Gene Polymorphisms and Differential Susceptibility to Severe Malaria in Tanzania. *J. Infect. Dis.* *212*, 1129–1139.
 33. Ko, W.Y., Kaercher, K.A., Giombini, E., Marcatili, P., Froment, A., Ibrahim, M., Lema, G., Nyambo, T.B., Omar, S.A., Wambebe, C., et al. (2011). Effects of natural selection and gene conversion on the evolution of human glycophorins coding for MNS blood polymorphisms in malaria-endemic African populations. *Am. J. Hum. Genet.* *88*, 741–754.
 34. Mayer, D.C.G., Cofie, J., Jiang, L., Hartl, D.L., Tracy, E., Kabat, J., Mendoza, L.H., and Miller, L.H. (2009). Glycophorin B is the erythrocyte receptor of Plasmodium falciparum erythrocyte-binding ligand, EBL-1. *Proc. Natl. Acad. Sci.* *106*, 5348–5352.
 35. Wang, H.Y., Tang, H., Shen, C.K.J., and Wu, C.I. (2003). Rapidly Evolving Genes in Human. I. The Glycophorins and Their Possible Role in Evading Malaria Parasites. *Mol. Biol. Evol.* *20*, 1795–1804.
 36. Leary, S.C. (2010). Redox regulation of SCO protein function: controlling copper at a mitochondrial crossroad. *Antioxidants Redox Signal.* *13*, 1403–1416.
 37. Valnot, I., Osmond, S., Gigarel, N., Mehaye, B., Amiel, J., Cormier-Daire, V., Munnich, A., Bonnefont, J.P., Rustin, P., and Rötig, A. (2000). Mutations of the SCO1 gene in mitochondrial cytochrome c oxidase deficiency with neonatal-onset hepatic failure and encephalopathy. *Am. J. Hum. Genet.* *67*, 1104–1109.
 38. Miller, L.H., Baruch, D.I., Marsh, K., and Doumbo, O.K. (2002). The pathogenic basis of malaria. *Nature* *415*, 673–679.
 39. Martelle, S.E., Raffield, L.M., Palmer, N.D., Cox, A.J., Freedman, B.I., Hugenschmidt, C.E., Williamson, J.D., and Bowden, D.W. (2016). Dopamine pathway gene variants may

- modulate cognitive performance in the DHS - Mind Study. *Brain Behav.* 6, 1–12.
40. Zhou, Z., Yang, J., Wang, L., Zhang, H., Gao, Y., Shi, X., Wang, M., Kong, P., Qiu, L., and Song, L. (2011). A dopa decarboxylase modulating the immune response of scallop *Chlamys farreri*. *PLoS One* 6, e18596.
41. Menge, D.M., Zhong, D., Guda, T., Gouagna, L., Githure, J., Beier, J., and Yan, G. (2006). Quantitative trait loci controlling refractoriness to *Plasmodium falciparum* in natural *Anopheles gambiae* mosquitoes from a malaria-endemic region in western Kenya. *Genetics* 173, 235–241.
42. Steed, E., Rodrigues, N.T.L., Balda, M.S., and Matter, K. (2009). Identification of MarvelD3 as a tight junction-associated transmembrane protein of the occludin family. *BMC Cell Biol.* 10, 1–14.
43. Huang, D., Cancilla, M.R., and Morahan, G. (2000). Complete primary structure, chromosomal localisation, and definition of polymorphisms of the gene encoding the human interleukin-12 p40 subunit. *Genes Immun.* 1, 515–520.
44. Luty, A.J.F., Perkins, D.J., Lell, B., Schmidt-Ott, R., Lehman, L.G., Luckner, D., Greve, B., Matousek, P., Herbich, K., Schmid, D., et al. (2000). Low interleukin-12 activity in severe *Plasmodium falciparum* malaria. *Infect. Immun.* 68, 3909–3915.
45. Duvallet, E., Semerano, L., Assier, E., Falgarone, G., and Boissier, M.C. (2011). Interleukin-23: A key cytokine in inflammatory diseases. *Ann. Med.* 43, 503–511.
46. Minke, B., Wu, C.-F., and Pak, W. (1975). Genes required for cytotoxicity against virus-infected target cells in *K* and *D* regions of H-2 complex. *Nature* 254, 84–87.
47. Hill, A.V., Allsopp, C.E., Kwiatkowski, D., Anstey, N.M., Twumasi, P., Rowe, P.A., Benett, S., Brewster, D., McMichael, A.J., Greenwood, B.M., et al. (1991). Common west African HLA antigens are associated with protection from severe malaria. *Nature* 351, 595–600.
48. Chimusa, E.R., Mbiyavanga, M., Mazandu, G.K., and Mulder, N.J. (2015). AncGWAS: A post genome-wide association study method for interaction, pathway and ancestry analysis in homogeneous and admixed populations. *Bioinformatics* 32, 549–556.
49. Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M., and Price, A.L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* 46, 100–106.
50. Daya, M., Der Merwe, L., Galal, U., Möller, M., Salie, M., Chimusa, E.R., Galanter, J.M., Van Helden, P.D., Henn, B.M., Gignoux, C.R., et al. (2013). A panel of ancestry informative markers for the complex five-way admixed South African Coloured population. *PLoS One* 8, e82224.
51. Campbell, M.C., and Tishkoff, S.A. (2008). African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. *Annu. Rev. Genomics Hum. Genet.* 9, 403–433.
52. Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O., Choudhury, A., Ritchie, G.R.S., Xue, Y., Nsubuga, R.N., Young, E.H., et al. (2015). The African Genome Variation Project shapes medical genetics in Africa. *Nature* 517, 327–332.
53. Slatkin, M. (2008). Linkage disequilibrium - Understanding the evolutionary past and mapping the medical future. *Nat Rev Genet.* 9, 477–485.
54. Conrad, D.F., Jakobsson, M., Coop, G., Wen, X., Wall, J.D., Rosenberg, N.A., and Pritchard, J.K. (2006). A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat. Genet.* 38, 1251–1260.
55. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* 467, 52–58.
56. The 1000 Genomes Project Consortium (2011). Europe PMC Funders Group A map of human genome variation from population scale sequencing. *Nature* 467, 1061–1073.
57. Johnston, H.R., Hu, Y.J., Gao, J., O'Connor, T.D., Abecasis, G.R., Wojcik, G.L., Gignoux, C.R., Gourraud, P.A., Lizee, A., Hansen, M., et al. (2017). Identifying tagging SNPs for African specific genetic variation from the African Diaspora Genome. *Sci. Rep.* 7, 1–9.
58. Pe'er, I., Yelensky, R., Altshuler, D., and Daly, M.J. (2008). Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* 32, 381–385.
59. Popejoy, A.B., and Fullerton, S.M., (2016). Genomics is failing on diversity. *Nature*. 538, 161–164.
60. Piel, F.B., Howes, R.E., Patil, A.P., Nyangiri, O.A., Gething, P.W., Bhatt, S., Williams,

- T.N., Weatherall, D.J., and Hay, S.I. (2013). The distribution of haemoglobin C and its prevalence in newborns in Africa. *Sci. Rep.* 3, 1–8.
61. Schork, N.J., Murray, S.S., Frazer, K.A., and Topol, E.J. (2010). Common vs. Rare Allele Hypotheses for Complex Diseases. *Curr Opin Genet Dev* 19, 212–219.
62. Zuk, O., Hechter, E., Sunyaev, S.R., and Lander, E.S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci.* 109, 1193–1198.
63. Hall, M.D., and Ebert, D. (2013). The genetics of infectious disease susceptibility: Has the evidence for epistasis been overestimated? *BMC Biol.* 11, 1–5.
64. Karlsson, E.K., Kwiatkowski, D.P., and Sabeti, P.C. (2014). Natural selection and infectious disease in human populations. *Nat. Rev. Genet.* 15, 379–393.
65. Wilfert, L., and Schmid-Hempel, P. (2008). The genetic architecture of susceptibility to parasites. *BMC Evol. Biol.* 8, 1–8.
66. Mitchell-Olds, T., Willis, J.H., and Goldstein, D.B. (2007). Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nat. Rev. Genet.* 8, 845–856.
67. Dudbridge, F. (2016). Polygenic Epidemiology. *Genet. Epidemiol.* 40, 268–272.
68. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42, 565–569.
69. International Schizophrenia Consortium (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460, 748–752.
70. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82.
71. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* 81, 559–575.
72. Loh, P., Tucker, G., Bulik-sullivan, B.K., Vilhjálmsson, B.J., Finucane, H.K., Salem, R.M., Chasman, D.I., Ridker, P.M., Neale, B.M., Berger, B., et al. (2015). Efficient Bayesian mixed-model analysis increases association power in large cohorts.
73. Erbe, M., Hayes, B.J., Matukumalli, L.K., Goswami, S., Bowman, P.J., Reich, C.M., Mason, B.A., and Goddard, M.E. (2014). Erratum to “Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels” (*J. Dairy Sci.* 95:4114–4129). *J. Dairy Sci.* 97, 6622.
74. Speed, D., Hemani, G., Johnson, M.R., and Balding, D.J. (2012). Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.* 91, 1011–1021.
75. Speed, D., Cai, N., Johnson, M.R., Nejentsev, S., and Balding, D.J. (2017). Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* 49, 986–992.
76. Golan, D., Lander, E.S., and Rosset, S. (2014). Measuring missing heritability: Inferring the contribution of common variants. *Proc. Natl. Acad. Sci.* 111, E5272–E5281.
77. Bulik-sullivan, B.K., Loh, P., Finucane, H.K., Ripke, S., Yang, J., Working, S., Consortium, G., Patterson, N., Daly, M.J., Price, A.L., et al. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* 1–7.
78. Frank, S.A. (1994). Recognition and Polymorphism in Host-Parasite Genetics. *Philos. Trans. Biol. Sci.* 346, 283–293.
79. Tian, C., Hromatka, B.S., Kiefer, A.K., Eriksson, N., Noble, S.M., Tung, J.Y., and Hinds, D.A. (2017). Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat. Commun.* 8, 1–13.
80. Wei, W.-H., Hemani, G., and Haley, C.S. (2014). Detecting epistasis in human complex traits.
81. Cordell, H.G. (2007). Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet.* 2009 June ; 10: 392–404.
82. Castillejo-López, C., Delgado-Vega, A.M., Wojcik, J., Kozyrev, S. V., Thavathiru, E., Wu, Y.Y., Sánchez, E., Pöhlmann, D., López-Egido, J.R., Fineschi, S., et al. (2012). Genetic and physical interaction of the B-cell systemic lupus erythematosus-associated genes BANK1 and BLK. *Ann. Rheum. Dis.* 71, 136–142.
83. Evans, D.M., Spencer, C.C.A., Pointon, J.J., Su, Z., Harvey, D., Kochan, G., Opperman,

- U., Dilthey, A., Pirinen, M., Stone, M.A., et al. (2011). Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat. Genet.* *43*, 761–767.
84. Strange, A., Capon, F., Spencer, C.C.A., Knight, J., Weale, M.E., Allen, M.H., Barton, A., Band, G., Bellenguez, C., Bergboer, J.G.M., et al. (2010). A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. *Nat. Genet.* *42*, 985–990.
85. Williams, T.N., Mwangi, T.W., Wambua, S., Peto, T.E.A., Weatherall, D.J., Gupta, S., Recker, M., Penman, B.S., Uyoga, S., Macharia, A., et al. Negative epistasis between the malaria-protective effects of $\alpha + -$ thalassemia and the sickle cell trait. *Nature Genet.* *37*, 1253–1257.
86. Atkinson, S.H., Uyoga, S.M., Nyatichi, E., Macharia, A.W., Nyutu, G., Ndila, C., Kwiatkowski, D.P., Rockett, K.A., and Williams, T.N. (2014). Epistasis between the haptoglobin common variant and $\alpha +$ thalassemia influences risk of severe malaria in Kenyan children. *Blood* *123*, 2008–2016.
87. Opi, D.H., Swann, O., Macharia, A., Uyoga, S., Band, G., Ndila, C.M., Harrison, E.M., Thera, M.A., Kone, A.K., Diallo, D.A., et al. (2018). Two complement receptor one alleles have opposing associations with cerebral malaria and interact with $\alpha +$ thalassaemia. *Elife* *7*, 1–27.
88. van der Sijde, M.R., Ng, A., and Fu, J. (2014). Systems genetics: From GWAS to disease pathways. *Biochim. Biophys. Acta - Mol. Basis Dis.* *1842*, 1903–1909.
89. Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). Leading Edge Perspective An Expanded View of Complex Traits. *Cell* *169*, 1177–1186.
90. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic Localization of Common. *Science* (80-.). *337*, 1190–1195.
91. Spain, S.L., and Barrett, J.C. (2015). Strategies for fine-mapping complex traits. *Hum. Mol. Genet.* *24*, 111–119.
92. Chen, Z.J., Zhao, H., He, L., Shi, Y., Qin, Y., Shi, Y., Li, Z., You, L., Zhao, J., Liu, J., et al. (2011). Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. *Nat. Genet.* *43*, 55–59.
93. Schaid, D.J., Chen, W., and Larson, N.B. (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* *19*, 491–504.
94. Galarneau, G., Palmer, C.D., Sankaran, V.G., Orkin, S.H., Hirschhorn, J.N., and Lettre, G. (2010). Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. *Nat. Genet.* *42*, 1049–1051.
95. Farh, K.K.H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shores, N., Whitton, H., Ryan, R.J.H., Shishkin, A.A., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* *518*, 337–343.
96. Vaughan, A.M., Aly, A.S., Kappe, S.H.I., (2008). Malaria parasite pre-erythrocytic stage infection: Gliding and Hiding. *Cell Host Microbe*. *4*, 209–218.
97. Wang, L., Zhang, B., Wolfinger, R.D., and Chen, X. (2008). An integrated approach for the analysis of biological pathways using mixed models. *PLoS Genet.* *4*, e1000115.
98. Zeggini, E., and Ioannidis, J.P.A. Meta-analysis in genome-wide association studies (2009). *Pharmacogenomics*. *10*, 191–201.
99. Liu, C., Bousman, C.A., Pantelis, C., Skafidas, E., Zhang, D., Yue, W., and Everall, I.P. (2017). Pathway-wide association study identifies five shared pathways associated with schizophrenia in three ancestral distinct populations. *Transl. Psychiatry* *7*, e1037.
100. O'Dushlaine, C., Kenny, E., Heron, E., Donohoe, G., Gill, M., Morris, D.W., Corvin, A., O'Dushlaine, C.T., Quinn, E.M., O'Donovan, M.C., et al. (2011). Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. *Mol. Psychiatry* *16*, 286–292.
101. Kao, P.Y.P., Leung, K.H., Chan, L.W.C., Yip, S.P., and Yap, M.K.H. (2017). Pathway analysis of complex diseases for GWAS, extending to consider rare variants, multi-omics and interactions. *Biochim. Biophys. Acta - Gen. Subj.* *1861*, 335–353.
102. Griffiths, M.J., Shafi, M.J., Popper, S.J., Hemingway, C.A., Kortok, M.M., Wathen, A., Rockett, K.A., Mott, R., Levin, M., Newton, C.R., et al. (2005). Genomewide analysis of the host response to malaria in Kenyan children. *J Infect Dis* *191*, 1599–1611.

103. Lamonte, G., Philip, N., Reardon, J., Lacsina, J.R., Chapman, L., Thornburg, C.D., Telen, M.J., Ohler, U., Nicchitta, C. V, Haystead, T., et al. (2013). contributes to malaria resistance. *12*, 187–199.
104. Hasin, Y., Seldin, M., and Lusis, A. (2017). Multi-omics approaches to disease. *Genome Biol.* *18*, 1–15.